ChinaXiv合作期刊

空间科学大数据的机遇与挑战

邹自明* 胡晓彦 熊森林

中国科学院国家空间科学中心 北京 100190

chinaXiv:202303.00689v1

摘要 随着世界各国对空间科学的日渐重视和空间探测技术的蓬勃发展，空间科学大数据时代已经来临。文章从数据规模增长、数据管理理念演进、研究范式转变、大数据技术与工具发展、智能应用萌芽和研究生态系统建设等角度阐述了当前空间科学大数据发展的主要趋势，并结合未来发展需求与国家战略规划布局，剖析了空间科学大数据面临的具体挑战和发展机遇。文章提出应全面发力促进科学数据共享共用，扩大知识创新与科技产出，引领空间科学发展的新时代。

关键词 空间科学，科学大数据，规划建议

DOI 10.16418/j.issn.1000-3045.2018.08.017

空间科学覆盖宏观和微观两大领域，聚焦宇宙和生命的起源与演化、太阳系与人类的关系等前沿科学主题，着力解决宇宙演化、暗物质与暗能量、黑洞、引力波、太阳活动与空间天气、地球全球变化、地外生命形成与演化等困扰人类的科学问题，一般涉及空间天文、太阳物理、空间物理、行星科学、微重力科学、空间地球科学、空间基础物理和空间生命科学等分支领域，是具有高度创新导向性和前沿交叉性的学科领域[1]。空间科学是一门基于数据的实验科学，以卫星等空间飞行器为平台，结合地面观测台站或大型地基观测网络，获取大量科学探测及实验数据，开展数据分析、物理建模、关联挖掘等以数据为主线的科学工作，以期突破人类对自然的认知，促进基础科学进步，推动颠覆性技术创新，引领战略性新兴产业发展，并为国家安全带来强力的科技支撑。

世界强国均对空间科学给予高度重视和大力支持，设计规划了一系列的空间科学发展计划，如美国的《国家空间天气战略》[2]和《全球探索路线图》[3]、欧洲的《宇宙憧憬 2015—2025》[4]、俄罗斯的《2030 年前航天活动发展战略》[5]和我国的《中国空间科学项目中长期发展规划研究报告（2010—2030）》①等。在这些战略规划指引下，大科学装置、“旗舰”计划、专项计划和国际合作计划等重大空间科学任务得以有效推进和实施，科学数据快速积累，科学成果不断涌现。

*通讯作者
资助项目：中国科学院战略性先导科技专项（XDA19020500）
修改稿收到日期：2018年8月7日
① 中国科学院空间科学项目中长期发展规划研究课题组2008年编制。

ChinaXiv合作期刊

chinaXiv:202303.00689v1

1 空间科学迎来大数据时代

1.1 数据体量爆炸式增长

在空间科学发展规划的牵引下，空间科学探测正形成多波段、多信使、链网式、天地一体化联合探测的新格局。地基方面，东半球空间环境地基综合监测子午链（简称“子午工程”）一期已建成并投入使用[6]，子午工程二期工程也已启动建设[7]；以子午工程为基础，多台站、链网式、多学科交叉协同监测国际子午圈计划也正在筹划中[8]。天基方面，中国科学院战略性先导科技专项（A类）“空间科学”（简称“空间科学先导专项”）成功了发射“悟空”“SJ-10”“墨子”和“慧眼”4颗科学卫星，成果斐然[9-13]。空间科学先导专项二期（2020—2025年）预计发射GECAM、SVOM、SMILE、ASO-S、EP、EXTP序列科学卫星。。

系列重大项目成功实施后所带来直接成果便是科学数据的快速积累，依托于大视场、高分辨率、高灵敏度等新型观测技术，空间科学数据采集速率正以指数形式增长。空间科学先导专项一期4颗卫星已积累的科学数据超过200 TB，数据种类2 000多种，数据产品数近200万，预计至任务结束时，总数据量将超过420 TB；与一期相比，二期卫星任务有效载荷数据量以倍数提升，日均生产总数据量将超过8 000 GB，到二期任务末期，数据总量将达52.2 PB。

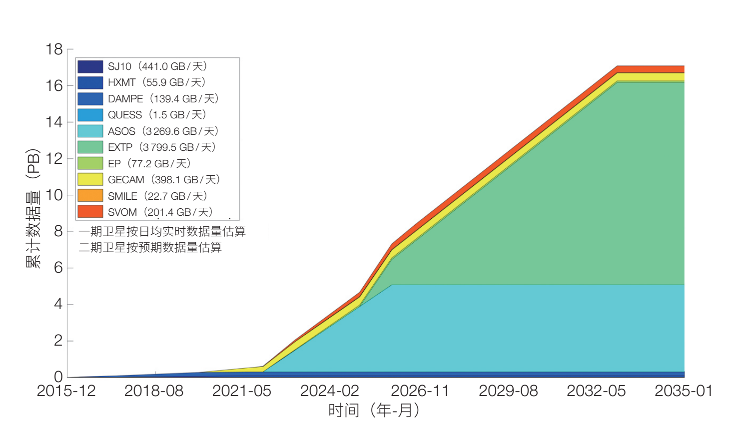


图1　空间先导一期和二期数据量估算

一期包含4颗卫星：DAMPE表示“悟空”，SJ10表示“SJ-10”，QUESS表示“墨子”，HXMT表示“慧眼”。二期包含6颗卫星：GECAM、SVOM、SMILE、ASO-S、EP、EXTP。本图数据均不考虑卫星延长服役的情况

美国国家航空航天局（NASA）空间科学卫星编目②显示，自2000年至今，世界范围发射的空间科学卫星达到674颗，年平均数量超过35颗。而在2016和2017年，更是以每年超过2倍的数量在增长（图2）。综合国内外空间科学天、地基平台的观测数据，空间科学年均数据生产率超过EB量级，从数据规模和体量角度来看，空间科学迎来了大数据时代。

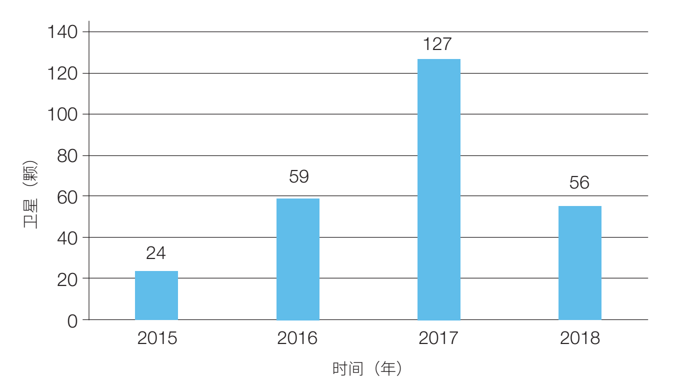


图2 2015—2018年世界范围空间科学卫星发展趋势

1.2 数据管理与保存得到重视

重大项目计划产生的海量科学数据作为国家资源和人类知识库，应进行长期保存和管理，让科学家在未来很长一段时间内可以深入挖掘数据背后的知识。国内外纷纷建立空间科学领域的数据中心/系统，实施数据的长期保存和质量管控。

美国国家航空航天局（NASA）建立了国家空间科学数据中心（National Space Science Data Center，NSSDC）用以保障空间科学卫星任务（含空间天文、天文物理、太阳和空间等离子体物理、行星科学、月球科学和空间物理）档案数据永久安全和长期可用，并为世界范围的科学家提供公开数据服务。截至2015年12月，仅其所属的行星数据系统（Planetary Data System，PDS）存储和管

② NASA空间科学卫星编目：https://nssdc.gsfc.nasa.gov/nmc/SpacecraftQuery.jsp。

ChinaXiv合作期刊

理的火星、月球、金星、水星各等级轨道探测数据已超过947 TB。

欧洲空间天文中心（European Space Astronomy Centre，ESAC）作为欧空局（European Space Agency，ESA）空间科学数据中心，也对所有欧洲空间天文、太阳系探测、行星科学、基础物理等卫星任务产生的科学档案数据进行了集中管理和存储。

在空间科学先导专项支持下，我国也建设了集科学卫星数据汇集与分发，数据全生命周期质量控制，数据存储、管理与归档为一体的空间科学先导数据管理系统（空间科学数据中心），有效支撑了“悟空”“SJ-10”“墨子”和“慧眼”卫星任务的实施，促进了卫星成果产出，可保障数据的永久安全。

1.3 研究范式发生转变

科学大数据时代的典型特征是研究范式的转变[14]，与传统基于少量数据样本开展理论分析的科研模式不同，大数据时代下的科研人员主要通过对多源、多要素、全样本空间的大数据进行分析，通常结合神经网络、机器学习等大数据技术，挖掘科学大数据中蕴藏的科学知识。空间科学领域研究模式也正向数据密集型科学发现模式转变。

作为数据驱动知识发现的典范，暗物质粒子探测卫星——“悟空”的科学家团队通过对卫星530天采集的28亿份高能宇宙射线数据样本分析，首次找到了电子宇宙线能谱在~1 TeV处的拐点（异常波动），而这个拐点反映着高能电子辐射源的典型加速能力，拐点下降行为对解释电子宇宙线是否来自暗物质起着关键作用[9]。

针对开普勒太空望远镜（Kepler space telescope）获取的海量数据，NASA科学家利用深度学习算法构建的机器学习模型具备对低信噪比数据进行自动系外行星识别能力，模型对开普勒太空望远镜数据库20万个目标星系数据进行自动搜寻，成功从中找到了Kepler-80 g和Kepler-90 i两颗系外行星[15]。

早在20世纪90年代，空间物理学研究中便开始采用机器学习等大数据技术对卫星获取的数据进行分析，开展空间天气研究和预报[16]。诸如磁层亚暴触发识别[17]，太阳活动（日冕物质抛射、耀斑）预测[18,19]和行星际激波预报[20]等。其中太阳耀斑预测[19]更是使用SDO卫星4年，超过5.5 TB的太阳光球层、色球层等图像大数据作为模型的训练输入。事实证明，大数据分析技术对非线性空间天气过程研究和高度复杂度空间天气事件预报具有重要实践意义，数据密集型的研究模式正逐渐发展成空间物理学的主流模式。

1.4 大数据技术和工具蓬勃发展

大数据分析技术、工具、算法和大数据系统平台是大数据研究和应用的关键环节，目前均有着长足的发展。

（1）算法成熟可用。除了上述提及的大数据技术和算法，学界也发展了很多其他大数据分析与挖掘算法，比如决策树、贝叶斯、神经网络、支持向量机等分类算法，*K*-means聚类、层次聚类、基于密度或基于网络的聚类等聚类算法，线性回归、逻辑回归等预测算法，以及卷积神经网络等适用于图像处理的机器学习模型等。这些算法在商业、城市治理等领域以及生物学、医药、地学等学科中均体现了巨大的潜力，同样也可被应用于空间科学研究领域。

（2）工具套件化、便捷化。美国分析图形公司（Analytical Graphics, Inc.，AGI）开发了“系统工具套件”（system took kit，STK），可为工程师和科学家提供四维建模、仿真、分析、操作，能对地面、海洋、空气、空间中的物体展开复杂的分析，为未来运行在这些环境中的系统或载荷的性能提供模拟计算和实时评估（图3）。比利时联邦科学研究开发的空间环境信息系统（space environment information system，SPENVIS）是空间环境与载荷、宇航员相互作用评估综合软件，集成了丰富的模式计算、仿真模型以及配套的数据分析工具。用户可通过Web访问SPENVIS构造模型、定义参数和执行仿真，实现研发、优化、跟踪和检测空间飞行器与探测器的性能、工作状态、故障，预测空间环境及其影响。

chinaXiv:202303.00689v1

ChinaXiv合作期刊

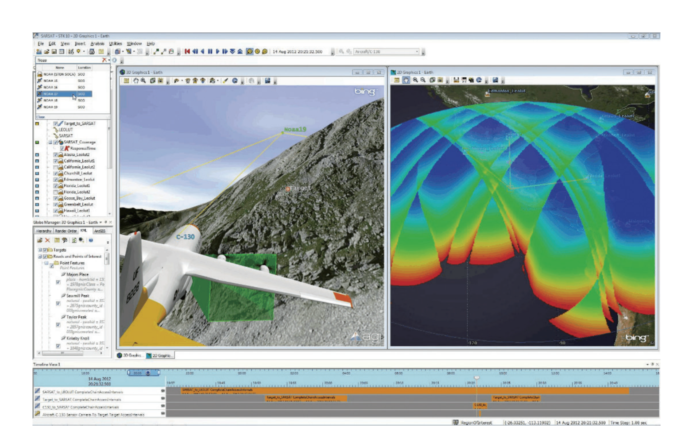


图 3 “系统工具套件”（STK）仿真示意图

（3）服务系统集成化、融合化。NASA 戈达德飞行中心（GSFC）建立的协同数据分析网（CDA Web）汇集了 1992 年至今国际上近乎所有卫星任务所产生的探测数据，以及部分地面站网观测数据，真正实现了一站式的科学数据查询、物理量抽取、可视化绘制、下载和在线分析功能，是空间科学领域世界上应用最广泛的数据系统。NASA 建立的网页建模目录和建模档案库（CCMC）集成了大量的可计算模型，如大气风场模型、电离层电子密度模型、太阳与行星际空间模型、星型（金星、火星、木星）模型。现有模型基本涵盖了整个空间物理领域的模式需求，可为用户提供模型在线计算和可视化分析，支持空间物理研究与空间天气预报。日地空间系统研究网络（STAR-Network）[21]是国内发展的集 IT 基础设施、多要素空间科学数据、数据分析工具套件和空间天气模式于一体的空间科学数据应用平台，具备空间科学任务支持及科研创新活动服务能力。

1.5 空间大数据智能应用趋势明显

面向智能应用的新科学概念正在形成。例如，魏奉思院士提出“数字空间”战略[22]，以空间科学天、地基观测数据为驱动，以科学认知为依据，以云计算基础设施及空间大数据应用技术为手段，打造集空间科学、空间技术、空间应用与空间服务为一体的重大空间基础设施，数字化呈现真实宇宙空间时空要素变化，开启应对空间天气灾害，增强卫星应用能力，服务开拓空间新能源、新通信、新交通、新制造、新环保等战略经济新领域。在新科学概念牵引下，高维时空离散技术[23]、智能融合数据挖掘与应用技术[24]等技术也取得了重要进展。

1.6 空间大数据研究与应用生态系统正在形成

在上述空间科学大数据发展趋势下，空间科学领域的大数据研究与应用生态系统逐渐形成。开源社区遍地开花，例如空间天文领域科学家组织了 Astropy、Openastronomy 等开源社区系统，也组建了天文圈、AstroIDL、AstroTex 等大量交流群组。学术交流活动呈现欣欣向荣的景象，有 BiDS（Big Data for Space）等空间科学领域专门的大数据研讨会，也有美国地球物理联合会（AGU）等国际顶级学术会议针对数据科学等方向设立的分会，我国自 2014 年开始组织的科学数据大会也为国内各学科大数据研究者们提供了交流、展示与研讨的机会。空间科学大数据研究网络已初具规模，国内外相关科研机构与高校的团队针对数据开放共享、数据管理保存、工具研发、智能应用、基础设施建设等研究热点开展了大量工作，并积极响应科学数据出版等倡议，共同为打造良好的学科大数据生态不懈努力。

2 空间科学大数据的挑战与发展机遇

空间科学大数据正呈现欣欣向荣的发展趋势，取得了一系列可观的研究与建设成果，为空间科学发展带来了新的契机。同时我们也需要看到，由于缺乏面向领域大数据发展的顶层规划，以及学科社区对大数据发展的认知和信心不足等原因，我国空间科学大数据发展也面临诸多瓶颈与挑战，总体呈现挑战与机遇并存态势，下面从 5 个方面阐述。

2.1 数据开放共享面临挑战与机遇

国内空间科学数据在机构间、项目间的流动性不足，普遍存在数据标准不一、数据质量参差不齐的问题，且缺少良好的数据共享机制与服务平台，阻碍了科研创新工作的进一步开展。正值国家出台《科学数据管理办法》的大好时机，应着力提升科学数据开放共享理念，从数据

chinaXiv:202303.00689v1

chinaXiv:202303.00689v1

开放性、规范性、安全性角度考虑，研究合适的空间科学数据标准与共享规范，实现空间科学数据的整合交汇与开放交流，促进高价值科学数据的充分共享和使用。

2.2 大数据基础设施面临挑战与机遇

与高能物理、生物医药等学科相比，目前国内空间科学领域的传输网络、计算资源、应用软件和算法工具等大数据基础设施发展相对滞后，没有长期规划，缺乏顶层设计，且基础能力相对薄弱、分散，导致国家对空间科学领域的支持力度不足。例如，特定的专业模型与算法向公共超算资源部署存在困难，传输网络问题（速率、限制）使科学数据访问、交换受到限制等。

因此，应重视大数据基础设施长期规划，加强顶层设计，合理布局建设，有针对性地增强空间科学大数据基础设施能力。例如，通过建设必要的国际数据传输网络、打造面向空间科学领域专业需求的公共计算环境等，匹配和缓解未来空间科学重大任务和创新研究活动的巨大压力。同时，通过政策引导以及增加虚拟化设施、增强虚拟化资源调度能力，使尽可能多的大数据基础设施实现联合共享。

2.3 数据长期安全面临挑战与机遇

目前，国内的空间科学数据资源主要由科研机构或项目部门各自保存，在日常管理中更重视数据使用的便捷性，而数据长期安全的保障能力相对不足。应充分结合重大任务开展数据活动过程域管理，实现数据活动全生命周期的规范化管理，通过全面的质量控制方案保障数据的科学性与可靠性，在此基础上充分利用先进的科学数据管理与灾备技术，保障科学数据的永久安全和长期可用。

2.4 颠覆性领域大数据管理、分析与应用技术亟待突破

数据密集型知识发现和智能融合应用平台建设对领域大数据管理、分析与应用技术需求急迫，但现状是传统科研团体对大数据助力科研产出缺乏重视，同时针对领域专业算法与软件投入力度较小，自主研发能力相对薄弱，导致颠覆性新技术积累不足。应加大在大科学任务、大科学装置中大数据技术研发与应用的投入，并借鉴美国 EarthCube、ROSES 等发展行动计划，在我国设立空间科学大数据技术研究基金，长期支持空间科学大数据先进技术的发展。

2.5 研究应用生态系统建设迎来新机遇

鉴于此，我们应积极响应国家号召，抓住空间科学发展的良好态势，利用好重大空间科学项目实施的关键契机，建立国家级空间科学数据中心，组建创新交叉人才队伍，共同发展空间科学大数据开源社区，使这些社区成为创新工场，不断为空间科学大数据发展提供新的思路。研发面向领域的专用算法与应用工具，弥补我国空间科学大数据自主创新能力不足的状况，促进空间科学大数据研究与应用生态良性发展。

3 结语

受益于党和国家的高度重视，中国空间科学事业发展迅猛，空间科学长远发展规划布局合理，空间科学迎来了大数据时代。大数据思潮下，空间科学发展呈现了崭新的局面，我们应抓住空间科学大数据发展的趋势与机遇，在数据开放共享、数据存储管理、大数据基础设施布局、大数据关键技术突破与研究应用生态系统建设等各个方面全面发力，将我国空间科学、空间技术与空间应用推向新的高度。

参考文献

1 吴季. 2016—2030年空间科学规划研究报告. 北京: 科学出版社, 2016.

2 Space Weather Operations, Research, and Mitigation (SWORM) Task Force Co-Chairs. National Space Weather Strategy. [2018-08-04]. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/final_nationalspaceweatherstrategy_20151028.pdf.

3 International Space Exploration Coordination Group. Global exploration roadmap. [2017-01-05]. https://www.nasa.gov/sites/

ChinaXiv合作期刊

chinaXiv:202303.00689v1

default/files/files/GER-2013_Small.pdf.

4 European Space Agency. Space science for Europe 2015-2025. [2017-01-05]. http://www.esa.int/esapub/br/br247/br247.pdf.

5 尤亮, 白青江, 孙丽琳, 等. 世界主要空间国家空间科学发展态势综述. 中国科学院院刊, 2015, 30(6): 740-750.

6 Wang C. New chains of space weather monitoring stations in China. Bulletin of the Chinese Academy of Sciences, 2013, 27(1): 35-40.

7 王赤, 任丽文. 日地空间探索之旅——空间物理探测最新进展与展望(下). 国际太空, 2015(02): 82-86.

8 范全林. 基于子午工程的国际空间天气子午圈计划. 中国科学基金, 2008(2): 65-69.

9 Collaboration D, Ambrosi G, An Q, et al. Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons. Nature, 2017, 552(7683): 63-66.

10 Liao S K, Cai W Q, Liu W Y, et al. Satellite-to-ground quantum key distribution. Nature, 2017, 549(7670): 43-49.

11 Liao S K, Yong H L, Liu C, et al. Long-distance free-space quantum key distribution in daylight towards inter-satellite communication. Nature Photonics, 2017, 11(8): 509-518.

12 Yin J, Cao Y, Li Y H, et al. Satellite-based entanglement distribution over 1200 kilometers. Science, 2017, 356(6343): 1180-1184.

13 Li T P, Xiong S L, Zhang S N, et al. Insight-HXMT observations of the first binary neutron star merger GW170817. Science China (Physics, Mechanics & Astronomy), 2018, 61(3): 031011.

14 Hey T. The fourth paradigm-data-intensive scientific discovery. Proceedings of the IEEE, 2011, 99(8): 1334-1337.

15 Shallue C J, Vanderburg A. Identifying exoplanets with deep learning: a five planet resonant chain around Kepler-80 and an eighth planet around Kepler-90. Astronomical Journal, 2017, 155(2): 94.

16 Camporeale E, Wing S, Johnson J R. Machine learning techniques for space weather. Amsterdam: Elsevier, 2018.

17 Sutcliffe P R. Substorm onset identification using neural networks and Pi2 pulsations. Annales Geophysicae, 1997, 15(10): 1257-1264.

18 Bobra M G, Ilonidis S. Predicting coronal mass ejections using machine learning methods. Astrophysical Journal, 2016, 821(2): 127.

19 Jonas E, Bobra M, Shankar V, et al. Flare prediction using photospheric and coronal image data. Solar Physics, 2018, 293(3): 48.

20 Vandegriff J, Wagstaff K, Ho G, et al. Forecasting space weather: Predicting interplanetary shocks using neural networks. Advances in Space Research, 2005, 36(12): 2323-2327.

21 邹自明, 佟继周, 熊森林, 等. 大数据时代空间科学领域的科研信息化实践与成果. 大数据, 2016, 2(6): 83-96.

22 魏奉思. “数字空间”是空间科技战略新高地. 河南科技, 2016, (19): 7-7.

23 康栋贺, 邹自明, 胡晓彦, 等. 支持时空耦合计算的HTM-ST日地空间系统数据组织模型. 地球信息科学学报, 2017, 19(6): 735-743.

24 王天真. 智能融合数据挖掘方法及其应用. 上海: 上海海事大学, 2006.

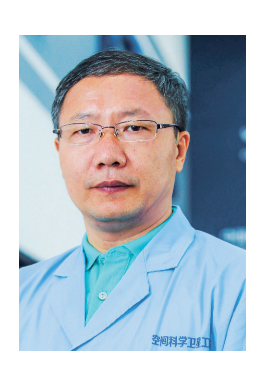
Challenges and Opportunities of Big Data in Space Science

ZOU Ziming* HU Xiaoyan XIONG Senlin
(National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China)

Abstract Space science is a discipline with high innovation orientation and frontier intersection. Countries all over the world attach great importance to it and have promoted a series of strategic planning and major programs. The age of big data in space science has arrived. In this study, the main trends of big data development in space science are expounded. Specifically, space scientific data volumes are exploding, data storage and management are valued, the scientific research paradigm is shifting, big data technology and tools are booming, the intelligent application is budding and a benign research ecosystem of big data has been formed. Based on the development requirements and national strategic planning, this study analyzes the specific challenges and development opportunities of big data in space science. An all-out efforts should be made, from the perspectives of data sharing, data long-term storage, big data infrastructure construction, disruptive technologies breakthrough and research ecosystem construction, to promote the open and sharing of scientific data, to expand intellectual innovation and scientific and technological output, and to create a new era for the development of space science.

Keywords space science, scientific big data, planning proposal

chinaXiv:202303.00689v1



邹自明 国家空间科学中心副主任、研究员，中国科学院空间科学战略性先导科技专项科学卫星工程地面支撑系统总指挥兼总设计师，世界数据系统（WDS）中国空间科学学科中心主任，中国地球物理学会信息技术专业委员会委员，中国科学院空间环境研究预报中心科技委员会委员。主要从事空间环境信息及模式系统集成，空间科学卫星数据处理，空间信息的组织、检索和互操作，日地空间信息表示与可视分析等研究。发表论文30余篇，合著专著1部。曾获军队科技进步奖一等奖两项、中国科学院载人航天工程重要贡献奖，被授予“中国科学院参加载人航天工程优秀工作者”荣誉称号。E-mail: mzou@nssc.ac.cn

ZOU Ziming Deputy Director of National Space Science Center, Chinese Academy of Sciences (CAS), the chief designer of the Ground Support System in CAS Strategic Priority Program on Space Science. He also serves as the director of Chinese Space Science Data Center in World Data System, member of Information Technology Committee in Chinese Geophysical Society, and member of Science and Technology Committee in CAS Space Environment Prediction Center. ZOU mainly engages in space environment information system integration and space science satellite data processing system construction, also develops research on space information such as solar-terrestrial information organization, retrieval, representation, visual analysis, interoperability, etc. Till now, he has published one book (co-authored) and more than 30 papers, achieved two first prize of Military Science and Technology Progress Award, CAS Manned Space Program Important Contribution Award, and honored as “the excellent worker in CAS Manned Space Program”. E-mail: mzou@nssc.ac.cn

■责任编辑：文彦杰

*Corresponding author